DataTrail Program Guides

September, 2023

Table of Contents

# About this Course

This course is part of a series of courses for [DataTrail](https://www.datatrail.org/). DataTrail is a no-cost, paid 14-week educational initiative for young-adult, high school and GED-graduates. DataTrail aims to equip members of underserved communities with the necessary skills and support required to work in the booming field of data science.

DataTrail is a fresh take on workforce development that focuses on training both Black, Indigenous, and other people of color (BIPOC) interested in the data science industry and their potential employers.

Offered by the Johns Hopkins Bloomberg School of Public Health, in partnership with local non-profits and Leanpub, DataTrail combines a mutually-intensive learning experience (MILE) with a whole-person ecosystem of support to allow aspiring data scientists and their employers to succeed.

DataTrail uses mutually-intensive learning DataTrail joins aspiring data science scholars and expert-level data scientist mentors in a mutually-intensive learning experience (MILE).

In the DataTrail MILE:

Scholars engage in cutting-edge technical and soft skills training needed to enter the data science field.

Mentors engage in anti-racism and mentorship training needed to be impactful mentors and informed colleagues on diverse data science teams.

The social connections created along the way will fuel job opportunities for scholars and foster a more diverse, equitable, and inclusive climate at the mentors’ institutions.

# 1 Background for DataTrail

DataTrail is a no-cost, paid educational initiative for young-adult, high school and GED-graduates. DataTrail aims to equip members of underserved communities with the necessary skills and support required to work in the booming field of data science.

DataTrail began as an initiative through the Johns Hopkins Bloomberg School of Public Health, in partnership with local non-profits in the Baltimore area. The materials, lessons, and strategies that the inaugural DataTrail program has used are described here in hopes that other groups in other cities can start their own DataTrail programs.

In this guide, we will cover how you can start your own DataTrail program.

## 1.1 Motivation

### 1.1.1 The bad news:

Data science suffers from a lack of diversity that likely stems from societal racial inequities absorbed by the field. Data science is ovewhelmingly and unfortunately white and male. Meaning the field is not equitable for entry for everyone.

This is a huge problem for two reasons:

1. Lack of diversity in data science work can leads to short-sighted and limited work that ultimately can be harmful to larger society:

* [How our data encodes systematic racism](https://www.technologyreview.com/2020/12/10/1013617/racism-data-science-artificial-intelligence-ai-opinion/) - Technology review
* [Racial bias in a medical algorithm favors white patients over sicker black patients](https://www.washingtonpost.com/health/2019/10/24/racial-bias-medical-algorithm-favors-white-patients-over-sicker-black-patients/) - Washinton Post
* [Many Facial-Recognition Systems Are Biased, Says U.S. Study](https://www.nytimes.com/2019/12/19/technology/facial-recognition-bias.html) - The New York Times
* [As Cameras Track Detroit’s Residents, a Debate Ensues Over Racial Bias](https://www.nytimes.com/2019/07/08/us/detroit-facial-recognition-cameras.html) - The New York Times
* [Facebook’s ad-serving algorithm discriminates by gender and race](https://www.technologyreview.com/2019/04/05/1175/facebook-algorithm-discriminates-ai-bias/) - MIT Technology Review
* [How community members in Ramsey County stopped a big-data plan from flagging students as at-risk](https://www.tcdailyplanet.net/how-community-members-in-ramsey-county-stopped-a-big-data-plan-from-flagging-students-as-at-risk/) - TC Daily Planet

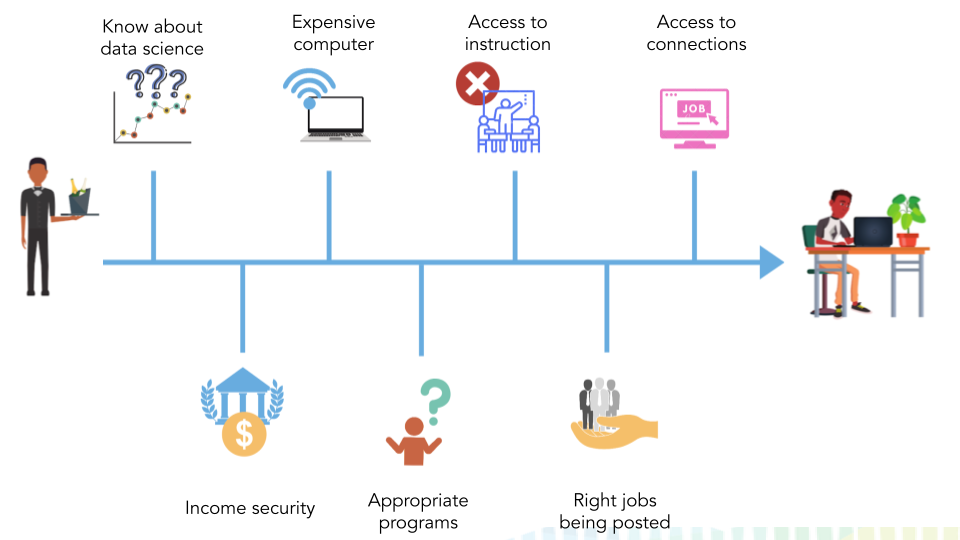
1. Income mobility is a public health issue and many individuals who would highly benefit from the career of data science are currently excluded from the benefits of an exciting and lucrative career in data science.

## 1.2 The barriers

Many industries have huge data science hiring demands but the individuals who could help fill this demand in ways that would bring sorely needed insights to the field are impeded by a number of barriers that can often be dispropotionately difficult barriers for marginalized individuals.

In order for any person to launch their data science career they need:

* Know data science is a career option!
* Have the income security that would allow them to pursue career building education
* Own a quality computer so they can do data science work
* Have connections and the ability to join the appropriate data science education programs
* Have access to data science instruction to learn the skills.
* Find appropriate job postings that fit their career goals
* Access to the connections needs to continue to build their data science career



### 1.2.1 The good news:

The goal of DataTrail is to try to reduce racial inequities in data science by reducing these barriers for young adults.

Income mobility is aided by education! So what we can do as a part of a DataTrail program is support these young individuals on their journey’s toward a data science career.



To this end, a DataTrail program provides:

* Payment to help with income security for individuals in the program
* A Chromebook laptop and access to Posit Cloud for use for data science work.
* Instruction and support for the data science education
* Social support to help scholars mitigate other barriers to their education: internet access, transportation issues, housing, childcare, etc.
* Internships after completion of the program
* Connections to individuals and programs that can help a DataTrail graduate build their career

In the upcoming chapters, we will discuss all the tips and strategies we have used to set up the Baltimore DataTrail program in hopes that you can use this guide run a DataTrail program.

# 2 Components of a program





## 2.1 Step 1: Non profit partners

A good DataTrail program needs to be rooted in a connection to the community it wishes to serve. The DataTrail program needs to respond to what the community wants and the needs they describe.

The role of the non-profit partner is to help with the social support and resources for the scholar’s development. For example, in the Baltimore, the inaugural cohorts were done in partnership with the Youth Opportunity program of the Baltimore Mayor’s office:

Youth Opportunity (YO) Baltimore serves youth between the ages of 18 and 24 who are disconnected from school and/or the workforce and links them to community-based educational, workforce preparation, and personal development services. Operating out of two safe, youth-friendly centers – one in West Baltimore and one in East Baltimore – YO embraces a youth development model that builds upon the existing strengths of each YO member, connects them with caring adults, and offers a full range of beneficial services.

A great non profit partner is one that has overlapping goals with DataTrail in a way that you can support each other. You may need to ask around in your community to find groups who are doing similar work.

### 2.1.1 Tips for finding non-profit partners:

TODO: I have no idea how this is done.

## 2.2 Step 2: Finding funding

You’ll need funding to start your own DataTrail program. Foundations, Philanthropic organizations, and Grant giving institutions can help you secure this funding.

A DataTrail costs include:

* A program administrators time
* A case manager’s time
* An lead instructor’s time
* Tutor(s) time - 2 students per tutor
* Chromebook per student (~$650)
* Stipends per student during the instruction (~$4200)

### 2.2.1 Tips for securing funding:

TODO: I have no idea how this is done.

## 2.3 Step 3: Finding staff

Finding a passionate and reliable staff to run a DataTrail program is a key step to ensuring success. Ideally all the people you include on your DataTrail team have a key understanding of the goals and motivation of this program.

The number of hours that each of the positions are needed for will be dependent on the number of scholars that you run each year (and this will likely be determined by funding availability). If you run one small cohort a year, then all of these positions could be part-time or paid effort percentages of someone’s time in an academic setting. However, if you secure the funding to do so, you could definitely increase the hours of each of these team members to the point that these roles could be full time positions year round.

It is advisable that these individuals generally be located in the same geographic location and for paperwork purposes, employed by the same institution (cross institutional funding mechanisms may result in overhead fees and administrative headaches).

### 2.3.1 Program administrator

The program administrator ensures all the many pieces administration that are required to keep the DataTrail program are managed. This means managing finances as well as partnerships with community non profits and funders and internship locations.

General summary/purpose:  
  
The program administrator's role is to oversee the program components and ensure their continued progress. It is their job to negotiate partnerships with the non profit community partners and coordinate and manage the DataTrail staff. They ensure that everyone is compensated: both staff and scholars in the program.  
  
Hours per week: 20 - 25  
  
Specific duties & responsibilities:   
  
- Maintain and manage the financial budget for DataTrail program.  
- Negotiate partnerships with the non profit community partners.  
- Manage relationships with internship locations for DataTrail scholars upon graduation  
- Assist with funding requests to philanthropic.  
- Coordinate and manage the DataTrail staff.  
- Set up and maintain payroll for the staff of DataTrail.  
- Set up and maintain the payroll for the scholars of the DataTrail program (this will likely be coordinated through a non profit partnerships).  
- Experience with financial and other types administration.

### 2.3.2 Case Manager

The case manager role is crucial for the support of the DataTrail scholars and mitigating any barriers that may arise. The case manager’s main responsibility is to help the scholars look ahead to see what barriers they may encounter and cooperatively figure out a plan for how the scholar can combat these barriers.

The case manager has a working relationship with the lead instructor in regards to office hours attendance and negotiating reasonable accommodations for students in regard to their course completion and assignments if/when a situation calls for it.

General summary/purpose:  
  
The case manager will maintain a caseload of DataTrail scholars at various stages of course completion (enrolled, at-risk, or reinstated), as well as scholars on leave, and program alumni in an effort to stabilize retention and proactively anticipate and address the needs of those with concerns of physical and/or mental health, housing and food security, finances, and social adjustment that may affect their performance and other aspects of their DataTrail experience. The case manager will utilize a variety of interventions, referrals and follow-up services that best address scholar and alumni needs.  
  
This position will serve as the primary resource for scholars who are on leave from DataTrail by maintaining regular communication and subsequently offering support, resources and mentoring upon the scholar’s return to DataTrail. While strong, supportive relationships may develop with scholars, the case manager does not engage in individual therapeutic relationships.  
  
Hours per week: 15 - 20  
  
Specific duties & responsibilities:   
  
- Conduct intake assessments with each incoming scholar to highlight potential areas of concern and develop an individual success plan based on their unique circumstances. Meet 1:1 thereafter to track progress towards their personal, academic and professional goals.  
- Collect, document and evaluate details of scholars’ cases as reported by the scholar, DataTrail tutors, program manager and [Non profit partner].  
- Formulate intervention plans in order to enhance the academic and professional success of scholars.  
- Contact and collaborate with other YO and other community agencies to expedite resolution of problems.  
- Provide referrals to support services with YO and other community agencies.  
- Schedule regular follow-up communications with scholars on leave to determine if and when they are fit to return.  
- Facilitate reinstatement of any scholars on leave who wish to return to DataTrail.  
- Develop and implement an action plan to stabilize and increase scholar retention rates per enrolled cohort.

### 2.3.3 Lead instructor

This person likely should have experience in the field of data science and an interest in education. Their main responsibilities are leading the office hours that are held each week and they may also assist in the one on one tutoring sessions for a subset of the scholars.

The lead instructor should have a working relationship with the program administrator, case manager, and tutors to make sure that the students are supported scholastically. Their job is to take the temperature of the cohort as it pertains to the curriculum. The office hours sessions may include a short group interactive activity but mostly open time for scholars to ask questions that may have arisen from their individual work throughout the week.

The lead instructor and the tutors both may answer asynchronous questions from the students as well.

General summary/purpose:  
  
This individual will take a lead and active role in teaching within the DataTrail program https://www.datatrail.org/ Their primary responsibility will be ensuring the scholastic progress of the DataTrail scholars. We are looking for an individual with extensive teaching experience with young professionals like those in our program. Must have experience with R and some knowledge of data science.  
  
Hours per week: 10 - 15  
  
Specific duties & responsibilities:  
  
- Teach data science in the form of office hours and lectures for our DataTrail students.  
- Host weekly one on one tutoring sessions with our students.  
- Coordinate with the program administrator and tutors to help coordinate the support the DataTrail students  
- Attend weekly DataTrail team meetings  
- Willing to build up programming/data science knowledge as needed for teaching our DataTrail scholars.  
- Familiarity with R required.  
- Familiarity with GitHub preferred.

### 2.3.4 Tutors

Tutors should be individuals with data science field experience. It may be a good strategy to look for graduate students in statistics or computer science who know the basics in data science. The tutors’ responsibilities are to assist the lead instructor for making sure scholars are on track and assist with questions and troubleshooting. Tutors and the lead instructor each partner with a subset of the students and meet with them 1:1 to assist with their data science questions and foster an open line for communication for their data science topics or career-related questions.

General summary/purpose:  
  
This contracted position will take an active role in teaching within the DataTrail program https://www.datatrail.org/ We are looking for an individual with extensive teaching experience with young professionals like those in our program. Must have experience with R and some knowledge of data science.  
  
Hours per week: 10 - 15  
  
Specific duties & responsibilities:  
  
- Assist the teaching of the curriculum in conjunction with lead instructor in office hours each week.  
- Host weekly one on one tutoring sessions with our students.  
- Coordinate with the program administrator, lead instructor, and other tutors to help coordinate the support the DataTrail students  
- Willing to build up programming/data science knowledge as needed for teaching our DataTrail scholars.  
- Familiarity with R required.  
- Familiarity with GitHub preferred.

## 2.4 Step 4: Finding the scholars

This is something your Program Administrator and Case Manager will coordinate with your community non-profit partner.

TODO: Have Liz write about her process.

Additionally, there is an open

## 2.5 Step 5: Setting up to run your first cohort

## 2.6 Step 6: Internships placements

# 3 Setting up for a Cohort

The DataTrail program you and your team run should be customized to your student’s and staff’s needs and preferences. In this section we will describe suggestions for how a DataTrail program can be run, but it is totally up to you and your team which of these examples and advice you choose to follow or not.

Setting up a DataTrail cohort:

* ☐ Identify scholars that are invited to join this cohort.
* ☐ Set up a payroll mechanism for paying the DataTrail scholars as they complete the DataTrail curriculum
* ☐ Create a barrier management plan for each scholar
* ☐ Make expectations of the program clear. Have each scholar sign a DataTrail Program agreement [See template agreement here](https://docs.google.com/document/d/142K3rVvfgjjEnog-NgIlwvbYjdGGp8du/edit)

## 3.1 Identifying scholars

**Relevant Role:** Case Manager and/or non profit

Depending on the determined roles of your non-profit partnership, recruiting DataTrail scholars may or may not be one of your team’s responsibilities.

TODO: Liz knows these things. What tips does she have?

## 3.2 Payroll mechanisms

Ideally your non-profit partner or a non-academic institution can handle payroll in a relatively paperwork light way. It is critical that the payroll mechanism in place is truly ready to execute when the scholars have completed their first checkpoints. Recall that DataTrail is a paid program because income stability is often a barrier for entry into data science. This means scholars getting paid on time is critical not only for their stability but also their motivation. Furthermore, compensating individuals fairly and on time is something that is ethically important. Academia and other industries have often have us believe that people should be doing work because they enjoy it regardless of pay, but that is something that only income secure individuals can do.

TODO: Ashley and Liz know these things. What tips do they have?

## 3.3 Learning about your scholars

Relevant Role: Case Manager

TODO: Liz knows this stuff. TODO: Also Simone

In getting to know your scholars, you will need to discuss a few things with them:

### 3.3.1 Are they at a good place right now to commit to this educational endeavor?

### 3.3.2 Do they display a sincere interest in Data Science?

### 3.3.3 Is their housing and income stable enough for them to complete this program?

### 3.3.4 Do they understand the expectations and demands of this DataTrail program?

You should have some DataTrail program agreement that individuals sign to make this clear. [See template agreement here](https://docs.google.com/document/d/142K3rVvfgjjEnog-NgIlwvbYjdGGp8du/edit)

### 3.3.5 What kinds of barriers are they likely to encounter that may impede their ability to complete this program?

### 3.3.6 Will some of these barriers be mitigated or nonexistent at a later point in time?

### 3.3.7 What strategies and resources can the scholar employ to mitigate these barriers?

# 4 Running a DataTrail cohort

It’s a good idea to have a weekly team meeting amongst the case manager, instructor, tutors, and program administrator so any issues on any fronts that arise during the cohort can be discussed amongst the team.

Our assumption and advise is that if possible you run your cohort with office hours in person, however the Baltimore DataTrail program did run a cohort completely virtually at times.

## 4.1 Welcome week

The first week of the program should have the following goals:

* ☐ Scholars get to know staff and their fellow scholars – don’t underestimate how important human connection is!
* ☐ Set up expectations and motivations for the program
  + ☐ Signed DataTrail training agreements from each scholar. [See our template](https://docs.google.com/document/d/142K3rVvfgjjEnog-NgIlwvbYjdGGp8du/edit)
* ☐ Assess employment skills and make plan for preparing for work
* ☐ Assess in more detail scholar’s potential barriers in one on one’s
* ☐ Set up each scholar with a Chromebook (advisable to do software updates before handing these out).
* ☐ Set up payments payments, and calendar
* ☐ A Google Classroom set up for your cohort. [You can request that here](https://forms.gle/gACRsXNB3x3yZfLH6)
  + ☐ Invite all scholars, tutors, and the case manager to the GoogleClassroom

## 4.2 Components of the curriculum

**Relevant Roles:** Lead instructor, tutors, and Case Manager.

The course material is here: <https://datatrail-jhu.github.io/DataTrail/> (Except quizzes and swirl modules will not be present in this link).

**The course is split into 7 overall sections:**

* 00 Intro - (note this section doesn’t have a project!)
* 01 Forming Questions
* 02 Getting Data
* 03 Cleaning the Data
* 04 Plot the Data
* 05 Get the Stats
* 06 Share Results
* 07 Build your Resume - This project isn’t an Rmd but a website

For scholars to get credit and payment for work on each section they must:

1. Attend all office hours.
2. Completion and submission of associated projects
3. Completion and submission of associated quizzes
4. Completion and submission of associated swirl modules

## 4.3 Supporting students to meet requirements

It will be a partnership between the Case Manager and lead instructor and tutors to keep tabs on the scholars and figure out whether they have met the requirements for each section. If they haven’t met the requirements, that is something the Case manager and instructors will want to communicate and explore with the student about what type of support they may need to reach these requirements.

Unfortunately, not all DataTrail scholars will complete the program. Sometimes the reasons for this are life circumstances that impede their ability to complete the program. But other times, its merely that the scholar, with the help of the Case Manager, may determine that the DataTrail program is not the right fit for them at this time.

It should be a discussion between the scholar, the Case Manager, and when appropriate, the instructors about what the scholar needs most at this time to be supported. Sometimes, the best way to support the scholar is for them to discontinue the program OR attempt the program again at a later date if a particular circumstance might be improved.

TODO: This needs Liz and/or Simone to make this section better and more detailed.

## 4.4 Example schedule

A weekly schedule during a cohort for a given scholar may look like this:

| Sunday | Monday | Tuesday | Wednesday | Thursday | Friday | Saturday |
| --- | --- | --- | --- | --- | --- | --- |
|  | 2.5 hr individual work 30 min meeting with case manager | 2 hr individual work 1.5 hr for office hours | 2 hr individual work 1 hr for tutor one on one | 2 hr individual work 1.5 hr for office hours | 2 hr individual work |  |

Suggested structure for each office hour:

1. Start off by asking what questions folks have and if there’s anything from the material they’d like to go over as a group
2. Dive into a exercise – can use the ones below or something else if something comes up during the week that students have requested

| Session | Relevant course link | Suggested office hour material | Objectives |
| --- | --- | --- | --- |
| Welcome week | TODO: LINK TO COMPLETED CAREER READINESS | - Make sure they are on Slack and Google Classroom - Have them open up Zoom and practice sharing screens and show that they have access to the GoogleClassroom - Practice sending zoom chat as well | Comfort with the platforms and with Google Calendar |
| Office Hours 1 | [Intro 1](https://datatrail-jhu.github.io/DataTrail/intro-welcome-to-datatrail.html) | - Ice breakers and introductions. - Introduce what office hours will be like | - Making sure everyone is on Slack and Google Classroom and is [able to use their Chromebook](https://github.com/datatrail-jhu/instructor_resources/blob/master/01-chromebook.Rmd) - Try to make everyone comfortable with chatting – keep it informal and chill |
| Office Hours 2 | [Intro 2](https://datatrail-jhu.github.io/DataTrail/account-setup.html) | - Ask folks what data science is. - Address any lagging tech issues. | - Everyone should be on Basecamp and have an RStudio account |
| Office Hours 3 | [Forming Questions 1](https://datatrail-jhu.github.io/DataTrail/forming-questions.html) | - Discuss the steps of data science. - Walk through [a data science example](https://datatrail-jhu.github.io/DataTrail/the-data-science-process.html) - Discussing [How to Learn techniques](https://datatrail-jhu.github.io/DataTrail/how-to-learn.html) | Set up the mentality that data science is All about questions so questions are encouraged! AND that data science can be frustrating AND that is expected and okay! |
| Office Hours 4 | [Forming Questions 2](https://datatrail-jhu.github.io/DataTrail/forming-questions.html) | - Take a tour through RStudio and [demonstrate live how R Markdowns work](https://github.com/datatrail-jhu/office-hours-exercises/blob/main/Rmarkdown_practice.Rmd) - Cover the basics of [the first project in RStudio](https://datatrail-jhu.github.io/DataTrail/your-first-data-science-project.html) | - Discuss what questions can be answered by data science. - Get familiar with what an R Markdown looks like |
| Office Hours 5 | [Getting Data 1](https://datatrail-jhu.github.io/DataTrail/getting-data.html) | - [Going over R objects](https://alexslemonade.github.io/training-modules/intro-to-R-tidyverse/01-intro_to_base_R.nb.html) - Leave a lot of time for R questions. | - Asking how the first project went - Understanding R objects basics - [Cover debugging tips](https://datatrail-jhu.github.io/DataTrail/getting-help-in-r.html#debugging-code) |
| Office Hours 6 | [Getting Data 2](https://datatrail-jhu.github.io/DataTrail/getting-data.html) | - [Covering data frames exercise 1](https://github.com/datatrail-jhu/office-hours-exercises/blob/main/intro_r_exercises/intro_r_exercise_1.Rmd) - [Data frames exercise 2](https://github.com/datatrail-jhu/office-hours-exercises/blob/main/intro_r_exercises/intro_r_exercise_2.Rmd) | - Understand data frames - Leaving more time for R questions |
| Office Hours 7 | [Getting Data 3](https://datatrail-jhu.github.io/DataTrail/getting-data.html) | - [Cover file paths with RStudio demonstrations](https://github.com/datatrail-jhu/office-hours-exercises/blob/main/file_organization_exercises/file_organization_exercise.Rmd) - [Demonstrate R Markdown functionality](https://datatrail-jhu.github.io/DataTrail/r-markdown.html) | - Understand file paths - Be able to read in a CSV file - Be able to use an R Markdown - Be able to upload a file to RStudio - Practice loading libraries with library() |
| Office Hours 8 | [Getting Data 4](https://datatrail-jhu.github.io/DataTrail/getting-data.html) | [Preview Getting Data project](https://datatrail-jhu.github.io/DataTrail/getting-data-project.html) specifically [making sure everyone has made their Leanpub data googlesheet](https://datatrail-jhu.github.io/DataTrail/google-documents.html) | - Prep for Getting Data project - Look out for googlesheets credential issues - Practice using googlesheets4 |
| Office Hours 9 | [Cleaning the Data 1](https://datatrail-jhu.github.io/DataTrail/cleaning-the-data.html) | - Leave some time for wrapping up the Getting Data project - Start on [Tidying exercise 1](https://github.com/datatrail-jhu/office-hours-exercises/blob/main/data_tidying_exercises/data_tidying_exercise_1_unsolved.Rmd) | - Have Getting Data Project wrapped up - Get folks comfortable with the idea of tidy data - Show them about [TidyDataTutor](https://tidydatatutor.com/vis.html) |
| Office Hours 10 | [Cleaning the Data 2](https://datatrail-jhu.github.io/DataTrail/cleaning-the-data.html) | - Leave time for wrapping up Tidying exercise 1 - Can start [Tidying exercise 2](https://github.com/datatrail-jhu/office-hours-exercises/blob/main/data_tidying_exercises/data_tidying_exercise_2_unsolved.Rmd) | - Become more comfortable manipulating strings for data cleaning |
| Office Hours 11 | [Cleaning the Data 3](https://datatrail-jhu.github.io/DataTrail/cleaning-the-data.html) | - Can finish up [Tidying exercise 2](https://github.com/datatrail-jhu/office-hours-exercises/blob/main/data_tidying_exercises/data_tidying_exercise_2_unsolved.Rmd) - Can [introduce a tidytuesday case](https://github.com/rfordatascience/tidytuesday) if ahead of schedule | - Becoming more comfortable with cleaning data |
| Office Hours 12 | [Cleaning the Data 4](https://datatrail-jhu.github.io/DataTrail/cleaning-the-data.html) | - Introduce [Cleaning the Data Project](https://datatrail-jhu.github.io/DataTrail/cleaning-data-project.html) - As a group, walk [through the joining exercises in the material](https://datatrail-jhu.github.io/DataTrail/joining-data.html) | - Understand joins - Be ready to clean up a dataset from soup to nuts |
| Office Hours 13 | [Plotting the data 1](https://datatrail-jhu.github.io/DataTrail/plotting-the-data.html) | - Leave time for covering Cleaning Data Project wrap up - Cover [what makes a good plot](https://datatrail-jhu.github.io/DataTrail/good-plots.html) - [Start Data Viz exercise 1](https://github.com/datatrail-jhu/office-hours-exercises/blob/main/dataviz_exercises/dataviz_exercise_1_unsolved.Rmd) | - Be comfortable with the goal of data viz - Understand the basic formula of ggplots |
| Office Hours 14 | [Plotting the data 2](https://datatrail-jhu.github.io/DataTrail/plotting-the-data.html) | - [Start up data viz exercise 2](https://github.com/datatrail-jhu/office-hours-exercises/blob/main/dataviz_exercises/dataviz_exercise_2_unsolved.Rmd) | Becoming comfortable with ggplot2 |
| Office Hours 15 | [Plotting the data 3](https://datatrail-jhu.github.io/DataTrail/plotting-the-data.html) | - [Start up data viz exercise 3](https://github.com/datatrail-jhu/office-hours-exercises/blob/main/dataviz_exercises/Data_Visualization_Lab.Rmd) | Further becoming comfortable with ggplot2 |
| Office Hours 16 | [Plotting the data 4](https://datatrail-jhu.github.io/DataTrail/plotting-the-data.html) | - Wrap up any of the unfinished data viz exercises - Introduce [Plotting the data Project](https://datatrail-jhu.github.io/DataTrail/plotting-data-project.html) | - Being able to ask a question and then make a viz to answer it |
| Office Hours 17 | [Getting Statistics 1](https://datatrail-jhu.github.io/DataTrail/getting-statistics.html) | Introduce the concepts behind translating questions to stats Go through the [In practice chapter as a group](https://datatrail-jhu.github.io/DataTrail/in-practice-using-stats.html) | Set up the mentality for statistics and how one might use them in data science – emphasize that memorization is not needed! |
| Office Hours 18 | [Getting Statistics 2](https://datatrail-jhu.github.io/DataTrail/getting-statistics.html) | Discuss the difference between Descriptive, Exploratory and Inferential statistics by going through those chapters as a group | Understand the groups of kinds of statistical questions |
| Office Hours 19 | [Getting Statistics 3](https://datatrail-jhu.github.io/DataTrail/getting-statistics.html) | [Go through the “Playing with Stats” exercise](https://github.com/datatrail-jhu/office-hours-exercises/blob/main/stats_exercises/playing_with_stats.Rmd) | Try to get an intuitive sense for statistics and distributions |
| Office Hours 20 | [Getting Statistics 4](https://datatrail-jhu.github.io/DataTrail/getting-statistics.html) | Introduce the [Get the Stats Project](https://datatrail-jhu.github.io/DataTrail/stats-project.html) | Help students get comfortable with applying the stats |
| Office Hours 21 | [Sharing Results 1](https://datatrail-jhu.github.io/DataTrail/sharing-results.html) | Go through what version control means and why it aids in reproducibility. [Can use this exercise](https://github.com/datatrail-jhu/office-hours-exercises/blob/main/version_control_exercise/version_control.md). [Demonstrate how to link RStudio with GitHub](https://datatrail-jhu.github.io/DataTrail/cloning-a-repository.html) | - Help students learn about the importance of GitHub - Help students be able to version control their projects by linking to GitHub in RStudio |
| Office Hours 22 | [Sharing Results 2](https://datatrail-jhu.github.io/DataTrail/sharing-results.html) | [Demonstrate how to file a pull request](https://datatrail-jhu.github.io/DataTrail/pull-requests.html) and [cover GitHub terms and workflow using this chapter](https://jhudatascience.org/Adv_Reproducibility_in_Cancer_Informatics/using-version-control-with-github.html) | - Leave time open for GitHub terms - Emphasize that using version control and GitHub is a series of habits they can develop over time |
| Office Hours 23 | [Sharing Results 3](https://datatrail-jhu.github.io/DataTrail/sharing-results.html) | [Introduce the Final project](https://datatrail-jhu.github.io/DataTrail/github-and-final-data-project.html) and tell students to start thinking about what kind of data science question they’d like to ask and begin looking for datasets that fit. Ask Davon to show his example. | Encourage them to ask a data science question they are interested in! |
| Office Hours 24 | [Sharing Results 4](https://datatrail-jhu.github.io/DataTrail/sharing-results.html) | Help students brainstorm their projects and find data | Setting up students for their final project! |
| Office Hours 25 | [Sharing Results 5](https://datatrail-jhu.github.io/DataTrail/sharing-results.html) | Keep this office hours fairly open for students to ask for help with their final project, show their progress, ask for help where needed |  |
| Office Hours 26 | [Building a resume 1](https://datatrail-jhu.github.io/DataTrail/building-a-resume.html) | Encourage students to begin putting their final project in a presentation. Show example good presentations. Ask Davon to show his example presentation. | Prepare the students for presenting their project at graduation! |
| Office Hours 27 | [Building a resume 2](https://datatrail-jhu.github.io/DataTrail/building-a-resume.html) | Introduce the [Create your portfolio project](https://datatrail-jhu.github.io/DataTrail/create-your-portfolio.html) Have additional time for any help needed for their final project presentations | - Help the students prepare how to show off the work they’ve done - Leave time open for any questions about data science careers in general |
| Office Hours 28 | [Building a resume 3](https://datatrail-jhu.github.io/DataTrail/building-a-resume.html) | Have students take turns showing off their portfolio website! Wrap up any more help they need with their final project presentations | Celebrate the work they’ve done! |
| Graduation | Celebrate their amazing work! | Have the students celebrate the large amount of amazing work they’ve done in such a short amount of time – do not give critical feedback on the scholar presentations, these are all about encouragement and celebration! |  |

### 4.4.1 Wrapping up a cohort

As you are nearing the end of a cohort, you have likely learned a lot about your scholars. In this time, they’ve likely learned a lot about data science and are interested and ready for an internship which will allow them to further hone their skills!

TODO: How to do internship placements

### 4.4.2 To contribute to the curriculum:

The DataTrail curriculum is always looking to be improved. If you encounter issues, bugs, or otherwise find things in the curriculum that could use improvement, please let the curriculum developers know.

* You can email or Slack [csavonen@fredhutch.org](mailto:csavonen@fredhutch.org) with recommendations/problems/concerns.
* You can also post GitHub issues here: <https://github.com/datatrail-jhu/DataTrail/issues>
* All the associated DataTrail GitHub repositories are here: <https://github.com/datatrail-jhu> if you’d like to file pull requests.

\* Note that for self-learners (not a part of a cohort), [a Leanpub version of this material is available for certification here](https://leanpub.com/c/datatrail).

# 5 Internships

## 5.1 How to find internship locations

## 5.2 How to prepare internship locations

Not all workplaces are accustomed to younger individuals or what it means to mentor early career data scientists.

TODO: - Making expectations clear with internships - Giving guidance to mentors

## 5.3 How to pay for internships

## 5.4 How to place graduates in internships

## 5.5 How to support graduates in their internships

## 5.6 Ending Internships

## 5.7 Connecting scholars to job opportunities

# 6 Curriculum Development

# About the Authors

These credits are based on our [course contributors table guidelines](https://www.ottrproject.org/more_features.html#giving-credits-to-contributors).

| Credits | Names |
| --- | --- |
| **Pedagogy** |  |
| Lead Content Instructor(s) | [FirstName LastName](link%20to%20personal%20website) |
| Lecturer(s) (include chapter name/link in parentheses if only for specific chapters) - make new line if more than one chapter involved | Delivered the course in some way - video or audio |
| Content Author(s) (include chapter name/link in parentheses if only for specific chapters) - make new line if more than one chapter involved | If any other authors besides lead instructor |
| Content Contributor(s) (include section name/link in parentheses) - make new line if more than one section involved | Wrote less than a chapter |
| Content Editor(s)/Reviewer(s) | Checked your content |
| Content Director(s) | Helped guide the content direction |
| Content Consultants (include chapter name/link in parentheses or word “General”) - make new line if more than one chapter involved | Gave high level advice on content |
| Acknowledgments | Gave small assistance to content but not to the level of consulting |
| **Production** |  |
| Content Publisher(s) | Helped with publishing platform |
| Content Publishing Reviewer(s) | Reviewed overall content and aesthetics on publishing platform |
| **Technical** |  |
| Course Publishing Engineer(s) | Helped with the code for the technical aspects related to the specific course generation |
| Template Publishing Engineers | [Candace Savonen](https://www.cansavvy.com/), [Carrie Wright](https://carriewright11.github.io/), [Ava Hoffman](https://www.avahoffman.com/) |
| Publishing Maintenance Engineer | [Candace Savonen](https://www.cansavvy.com/) |
| Technical Publishing Stylists | [Carrie Wright](https://carriewright11.github.io/), [Ava Hoffman](https://www.avahoffman.com/), [Candace Savonen](https://www.cansavvy.com/) |
| Package Developers ([ottrpal](https://github.com/jhudsl/ottrpal)) [Candace Savonen](https://www.cansavvy.com/), [John Muschelli](https://johnmuschelli.com/), [Carrie Wright](https://carriewright11.github.io/) |  |
| **Art and Design** |  |
| Illustrator(s) | Created graphics for the course |
| Figure Artist(s) | Created figures/plots for course |
| Videographer(s) | Filmed videos |
| Videography Editor(s) | Edited film |
| Audiographer(s) | Recorded audio |
| Audiography Editor(s) | Edited audio recordings |
| **Funding** |  |
| Funder(s) | Institution/individual who funded course including grant number |
| Funding Staff | Staff members who help with funding |

## ─ Session info ───────────────────────────────────────────────────────────────  
## setting value   
## version R version 4.0.2 (2020-06-22)  
## os Ubuntu 20.04.5 LTS   
## system x86\_64, linux-gnu   
## ui X11   
## language (EN)   
## collate en\_US.UTF-8   
## ctype en\_US.UTF-8   
## tz Etc/UTC   
## date 2023-09-01   
##   
## ─ Packages ───────────────────────────────────────────────────────────────────  
## package \* version date lib source   
## assertthat 0.2.1 2019-03-21 [1] RSPM (R 4.0.5)   
## bookdown 0.24 2023-03-28 [1] Github (rstudio/bookdown@88bc4ea)   
## cachem 1.0.7 2023-02-24 [1] CRAN (R 4.0.2)   
## callr 3.5.0 2020-10-08 [1] RSPM (R 4.0.2)   
## cli 3.6.1 2023-03-23 [1] CRAN (R 4.0.2)   
## crayon 1.3.4 2017-09-16 [1] RSPM (R 4.0.0)   
## desc 1.2.0 2018-05-01 [1] RSPM (R 4.0.3)   
## devtools 2.3.2 2020-09-18 [1] RSPM (R 4.0.3)   
## digest 0.6.25 2020-02-23 [1] RSPM (R 4.0.0)   
## ellipsis 0.3.1 2020-05-15 [1] RSPM (R 4.0.3)   
## evaluate 0.20 2023-01-17 [1] CRAN (R 4.0.2)   
## fastmap 1.1.1 2023-02-24 [1] CRAN (R 4.0.2)   
## fs 1.5.0 2020-07-31 [1] RSPM (R 4.0.3)   
## glue 1.4.2 2020-08-27 [1] RSPM (R 4.0.5)   
## htmltools 0.5.5 2023-03-23 [1] CRAN (R 4.0.2)   
## knitr 1.33 2023-03-28 [1] Github (yihui/knitr@a1052d1)   
## magrittr 2.0.3 2022-03-30 [1] CRAN (R 4.0.2)   
## memoise 2.0.1 2021-11-26 [1] CRAN (R 4.0.2)   
## pkgbuild 1.1.0 2020-07-13 [1] RSPM (R 4.0.2)   
## pkgload 1.1.0 2020-05-29 [1] RSPM (R 4.0.3)   
## prettyunits 1.1.1 2020-01-24 [1] RSPM (R 4.0.3)   
## processx 3.4.4 2020-09-03 [1] RSPM (R 4.0.2)   
## ps 1.4.0 2020-10-07 [1] RSPM (R 4.0.2)   
## R6 2.4.1 2019-11-12 [1] RSPM (R 4.0.0)   
## remotes 2.2.0 2020-07-21 [1] RSPM (R 4.0.3)   
## rlang 1.1.0 2023-03-14 [1] CRAN (R 4.0.2)   
## rmarkdown 2.10 2023-03-28 [1] Github (rstudio/rmarkdown@02d3c25)  
## rprojroot 2.0.3 2022-04-02 [1] CRAN (R 4.0.2)   
## sessioninfo 1.1.1 2018-11-05 [1] RSPM (R 4.0.3)   
## stringi 1.5.3 2020-09-09 [1] RSPM (R 4.0.3)   
## stringr 1.4.0 2019-02-10 [1] RSPM (R 4.0.3)   
## testthat 3.0.1 2023-03-28 [1] Github (R-lib/testthat@e99155a)   
## usethis 1.6.3 2020-09-17 [1] RSPM (R 4.0.2)   
## withr 2.3.0 2020-09-22 [1] RSPM (R 4.0.2)   
## xfun 0.26 2023-03-28 [1] Github (yihui/xfun@74c2a66)   
## yaml 2.2.1 2020-02-01 [1] RSPM (R 4.0.3)   
##   
## [1] /usr/local/lib/R/site-library  
## [2] /usr/local/lib/R/library

# 7 References